**Nos. 23-11535-A, 23-11538-A, 23-11539-A**

---

**UNITED STATES COURT OF APPEALS**
**ELEVENTH CIRCUIT**

---

IN RE: DEEPWATER HORIZON BELO CASES

BELO PLAINTIFFS: LESTER JENKINS AND DWIGHT SIPLES,

Plaintiffs-Appellants,

*Versus*

BP EXPLORATION & PRODUCTION, INC., AND BP AMERICA PRODUCTION COMPANY,

Defendant-Appellees.

---

Appeal from U.S. District Court for the Northern District of Florida
Case Nos. 3:19-cv-00963, 5:19-cv-00260, 5:19-cv-00310

---

# AMICUS CURIAE BRIEF
**of Kenneth J. Rothman, DrPH**
**in Support of Plaintiffs-Appellants and Reversal**

---

**Benjamin James Stevenson**
Stevenson Legal, PLLC
919 Panferio Drive
Pensacola Beach, FL 32561
T. 702.306.6708
bjs@stevenson-legal.com

## CERTIFICATE OF INTERESTED PERSONS & CORPORATE DISCLOSURE STATEMENT

### Certificate of Interested Persons

The Certificate of Interested Persons contained in the Initial Brief is complete with the following additions:

1. **Rothman, Kenneth J., DrPH,** Amicus Curiae

2. **Stevenson Legal, PLLC**, counsel for Amicus Curiae Dr. Kenneth J. Rothman

3. **Stevenson, Benjamin James**, counsel for Amicus Curiae Dr. Kenneth J. Rothman

### Corporate Disclosure Statement

Not applicable. Amicus Rothman is a natural person and not a corporation or other business association.

## TABLE OF CONTENTS

## TABLE OF CITATIONS

Page

### Cases



### Rules



### Treatises

## IDENTITY & INTEREST OF AMICUS CURIAE

This case concerns the exclusion of expert testimony concluding contaminates from the Deepwater Horizon (DWH) spill may cause chronic sinusitis.

Amicus Curiae Kenneth J. Rothman is a Professor of Epidemiology at the Boston University School of Public Health. He is a Distinguished Fellow Emeritus at the Research Triangle Institute, an independent nonprofit research institute dedicated to improving the human condition. His research interests in epidemiology have spanned a range of health problems that includes cancer, cardiovascular disease, neurologic disease, birth defects, injuries, environmental exposures, and drug epidemiology. His main career focus, however, has been the development and teaching of the concepts and methods of epidemiologic research.

He was the founding editor of the journal *Epidemiology*, an Editor of the *American Journal of Epidemiology*, and a member of the Editorial Board of the *New England Journal of Medicine* and *The Lancet*. He has written two widely used epidemiologic textbooks: *Modern Epidemiology*, first published in 1986 and now in its fourth edition, and *Epidemiology – An Introduction*, now in its second edition. He is the recipient American College of Epidemiology's Lilienfeld Award for Excellence in Epidemiology, the Society for Epidemiologic Research's Career Accomplishment Award, and an honorary M.D. from the University of Aarhus. Although he rarely offers insights to courts, Dr. Rothman contributed to an amicus brief on

epidemiology to the U.S. Supreme Court. Brief Amici Curiae of Professor Kenneth Rothman et al., *Daubert v. Merrell Dow Pharmaceuticals, Inc.,* 509 U.S. 579 (1993), 1992 WL 12006438.

Dr. Rothman wants to ensure that jurisprudence tracks current science and epidemiology research. He submits this brief because this case involves principles of scientific inference that have been his career focus, and he wishes to discourage the use of misleading or inappropriate methodology, both in scientific discussions and in matters of societal import. The purpose of this brief is to demonstrate that the lower court opinion reflects a misunderstanding of the mechanics of scientific inference within the context of epidemiology.

All parties have consented to this filing. *See* Fed.R.App.P. 29(a)(2), 29(a)(4)(D).

**Disclosure Statement of Amicus Curiae**

The Plaintiffs' trial and appellate counsel, the Downs Law Group, will contribute money that is intended to fund the preparation and submission of his amicus brief. It will pay Dr. Rothman for his time preparing the brief and hiring legal counsel to assist him in the final preparation and submission of the brief.

No parties' counsel authored the brief in whole or in part.

No other person contributed money that was intended to fund preparing or submitting the brief.

## STATEMENT OF ISSUES

1. The plaintiffs' experts relied on literature that found an association between exposure to Deepwater Horizon containments and chronic sinusitis to guide their inference that the contaminants may cause chronic sinusitis. Yet, some findings in the literature were not labeled as "statistically significant." May the experts rely on a study's association as valid, even if specific finding of an association is not "statistically significant"?

2. The Bradford Hill factors are sometimes used to assess whether an agent-disease association is causal. Yet, their validity for this purpose is not established, nor is there a guide describing how to apply them. May an expert reliably find that the exposure to containments may cause chronic sinusitis without using these Bradford Hill factors?

## SUMMARY OF ARGUMENT

Federal Rule of Evidence 702 permits an expert to opine that contaminants may cause a disease, if (among other considerations) the testimony is the product of reliable principles and methods. The district court excluded the plaintiffs' experts' opinions because, in part, some results from a study on which the experts relied were not statistically significant, and because they failed to assess the Bradford Hill factors. This exclusion was incongruent with the science of epidemiology.

1. Statistical significance is a misleading criterion for assessing the validity or importance of associations, and its absence is not a proper basis to foreclose an inference. Experts reviewing a study need not reject an association because the study's author labeled a finding to be "not statistically significant."

2. An expert may conclude an association is causal without using the Bradford Hill factors. The best approach for reaching a causal conclusion derives instead from an assessment of evidence that indicates that competing non-causal theories cannot account fully for the observed association.

## ARGUMENT

**Scientists may reliably infer a causal relationship even when the association for one finding is not "statistically significant" and the scientist does not assess the Bradford Hill factors.**

A plaintiff may establish general causation in a toxic tort case through epidemiological evidence. *Chapman v. Procter & Gamble Distrib., LLC*, 766 F.3d 1296, 1308 (11th Cir. 2014). This method involves showing an association, often through literature, between the exposure to toxins and a disease. Then, an expert must determine whether the association reflects a true cause-effect relationship. Federal Judicial Center, REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 597 (Federal Judicial Center, 3d ed. 2011) ("Ref. Man.").

The district court determined the plaintiffs' experts failed to establish an association between Deepwater Horizon containments and chronic sinusitis. Order and Report and Recommendation (ECF 570) ("R&R") at 30-31, 55-65, 2022 WL 17721595.[1] The experts relied, in part, on the Rusiecki Study,[2] (ECF 543-1), which found a statistically significant association between oil inhalation and chronic sinusitis as reported in Table 4, ECF 543-1:7. R&R, ECF 570:29. The Rusiecki Study also

---

[1] *Report and recommendation adopted sub nom.* In re *Deepwater Horizon BELO Cases*, 3:19CV963, 2023 WL 2711573 (N.D. Fla. Mar. 30, 2023), Order (ECF 591), 2023 WL 2711573.

[2] Jennifer A. Rusiecki et al., *Incidence of Chronic Respiratory Conditions Among Oil Spill Responders: Five Years of Follow-up in the Deepwater Horizon Oil Spill Coast Guard Cohort Study*, 203 ENVIRONMENTAL RESEARCH 111824 (2022).

examined a subgroup and determined, although an association between oil inhalation and chronic sinusitis existed for never-smokers (Supplemental Table 5), that finding was not labeled as statistically significant. Rusiecki Study, ECF 543-1:23. In the district court's view, because the experts inadequately contended with the results from the never-smokers findings, this subgroup result rendered their reliance on other findings in the Rusiecki Study unreliable. R&R, ECF 570:31. The district court departed from how an epidemiologist should use statistical significance to assess associations. Epidemiologists understand that statistical significance depends not only on the strength of an association, but also on the amount of data, and therefore expecting subgroups of an overall population to show statistical significance belies an inadequate understanding of how statistical significance works.

Next, the district court faulted the experts for failing to meaningfully assess the Bradford Hill factors. R&R, ECF 570:32, 65. In its view, experts must use those factors to assess whether an association between exposure to oil inhalations and chronic sinusitis reflects a true cause-effect relationship. R&R, ECF 570:12. Because the district court found the experts failed to adequately use the Bradford Hill factors, it found experts' methodology unreliable and ruled inadmissible their causation opinions. *Id.* at 33, 60. Again, the district court departed from how an epidemiologist should establish a causal relationship. The strongest evidence for a causal conclusion

comes not from any checklist, but from an analysis that demonstrates that competing non-causal theories cannot account fully for the observed association.

Neither the strict adherence to statistical significance nor use of the Bradford Hill factors is appropriate, whether in pursuit of justice in a legal forum or for scientific inference in general.

**1. The association between oil inhalation and chronic sinusitis should not be discounted because one finding was not "statistically significant."**

Statistical significance is widely recognized as a misused, misleading analytic crutch that should be abandoned entirely. Blakeley B. McShane et al., *Abandon Statistical Significance*, 73 THE AM. STATISTICIAN 235 (2019).[3] Seven years ago, the American Statistical Association, in an official statement intended to discourage reliance on statistical significance testing, warned that:

> Scientific conclusions and business or policy decisions should not be based only on whether a *p*-value[4] passes a specific threshold. … A *p*-value, or statistical significance, does not measure the size of an effect or the importance of a result.

---

[3] Available at https://www.tandfonline.com/doi/full/10.1080/00031305.2018.1527253.

[4] A *p*-value measures "the probability of getting, just by chance, a test statistic as large as or larger than the observed value." Ref. Man. at 291. It informs the risk of random or sampling error.

RL Wasserstein et al, *The ASA Statement on p-Values: Context, Process, and Purpose*, 70:2 THE AM. STATISTICIAN 129, 131-132 (2016)[5]. The statement recognized that statistical significance has been widely taught and used because of a "circularity in the sociology of science based on the use of bright lines such as $p<0.05$: We teach it because it's what we do; we do it because it's what we teach." *Id.* at 129. The logical flaws in statistical significance testing and consequent mistaken inferences have been elaborated in many books and articles. *See, e.g.*, Sander Greenland et al., *Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations*. 31 EUR. J. EPIDEMIOL 337 (2016).[6] Indeed, the Reference Manual at 578-79, observes epidemiologists disagree about the appropriate role of statistical significance because other methods are available to guard against random or sampling errors:

> Epidemiologists have become increasingly sophisticated in addressing the issue of random error and examining the data from a study to ascertain what information they may provide about the relationship between an agent and a disease, without the necessity of rejecting all studies that are not statistically significant.

*Id.*

Characterizing a finding as either statistically significance or not statistically significant suffers from several fundamental problems:

---

[5] Available at https://www.tandfonline.com/doi/full/10.1080/00031305.2016.1154108

[6] Available at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4877414.

First,"statistically significant" does not mean that an association is strong, or causal. Similarly, "not statistically significant" does not mean that an association is absent, weak or not causal. If the strength of association is of interest, that is best measured directly, such as by the magnitude of a rate ratio, hazard ratio (HR), etc., rather than by labeling a finding as significant or not. Strong associations may be either statistically significant or not, and the same is true for weak associations, so statistical significance is not an effective way to distinguish large effects from small effects.

Second, statistical significance is typically evaluated by a statistic, the *p*-value, which is classified as significant if less than 0.05 and not significant if equal or greater than 0.05 (or equivalently, by determining whether a 95% confidence interval for the measure of association includes a relative risk of 1). A result with a *p*-value of 0.049 would be labeled as significant. If the *p*-value is 0.51, the test would be labeled as not significant. Yet, there is essentially no difference between these two *p*-values. The small difference between them implies that the interpretation of the two values should be essentially identical, not radically different. *See, e.g.*, Andrew Gelman and Hal Stern, *The Difference Between "Significant" and "Not Significant" is not Itself Statistically Significant*, 60 The Am. Statistician 328 (2012)[7] (these authors' lead sentence stated that "A common

[7] Available at https://www.tandfonline.com/doi/abs/10.1198/000313006X152649.

statistical error is to summarize comparisons by statistical significance and then draw a sharp distinction between significant and nonsignificant results.").

This issue applies directly to the Rusiecki Study. Rusiecki et al. found a statistically significant association between oil inhalation and chronic sinusitis. Rusiecki Study, ECF 543-1:7 (Table 4). When they focused on the subset of never-smokers, the association remained, and was in fact even stronger, but was no longer characterized as statistically significant. *Id.* at 23 (Supplemental Table 5). The results of the two findings are still consistent, despite one being labeled as statistically significant and the other not. The change in statistical significance comes from the subgroup being considerably smaller than the whole population; the number of observations in an analysis affects the determination of statistical significance.

Third, focus on statistical significance elevates chance above other factors that might account for an observed association. Chance is but one of many factors that can affect the magnitude of an observed association. Often, other more important factors should be considered, including selection bias, measurement errors, and confounding. For example, selection bias should be an overarching concern when patients receiving a promising drug fare worse than those given an older drug. It is not because the promising drug performs worse, but because it is prescribed for the sickest patients. Measurement error may be the dominant concern when measuring the health effects of alcohol drinking. The consequences of alcohol drinking may be

overestimated because when people are asked how much alcohol they consume, they typically underreport the actual amount. Thus, the effect of drinking two drinks per day might get recorded as the effect of drinking one drink per day, making alcohol consumption appear riskier than it actually is. Confounding, another common source of error, would be the primary concern in trying to understand why the mortality rate in Panama is lower than in Sweden. The mortality rate is not lower in Panama because Panama is a safer country, but because the population of Sweden is older on average than the population of Panama, and mortality increases with age. All these sources of error are more important considerations than chance in interpreting observed associations.

Fourth, some believe that statistical significance rules out chance as an explanation, but that is not the case. Suppose you flip a coin six times, getting 6 heads. Is this a fair coin, with equal probability of heads and tails? The *p*-value for 6 heads or 6 tails out of 6 flips is 0.031, a "statistically significant" result. According to the practice of those who rely on statistical significance for inference, one should conclude that the coin was not a fair coin. But chance is the most obvious explanation, a much more reasonable explanation than asserting that you are using a biased coin. Statistical significance testing does not rule out or rule in chance as an explanation for any finding. Judgment combined with other information is always required.

Because statistical significance testing is a flawed method, its use and acceptance in science is being challenged. Responsible scientists and others should heed the warnings of the American Statistical Association that the methodology of statistical significance testing is unreliable and misleading. This Court should do the same.

**2. Causal conclusions should derive from the scientific method, not strict adherence to the Bradford Hill factors.**

Based on the association between exposure to oil inhalation and chronic sinusitis found in the literature, the plaintiffs' experts considered and concluded the association reflects a true cause-effect relationship. Yet, the district court found they did not sufficiently evaluate the Bradford Hills factors. It excluded their expert testimony, in part, for this shortcoming. R&R, ECF 570:33, 66. This is incongruent with sound epidemiology.

The Bradford-Hill factors refer to a list of considerations to assess whether a causal inference may be made from an association. They were mentioned by the British statistician, Sir Austin Bradford Hill, in an after-dinner talk that was later published. He discussed nine considerations, reflected in the REFERENCE MANUAL at 600:

1.Strength of association
2. Consistency
3. Specificity
4. Temporality
5. Biological gradient
6. Plausibility
7. Coherence
8. Experimental evidence
9. Analogy

Although this informal list of considerations from Bradford Hill might be useful to ponder, it is in no way a checklist for causal inference. Indeed, Sir Bradford Hill himself warned against considering this list to be a list of criteria, a word that he studiously avoided. Rather, he claimed that:

> What I do *not* believe … is that we can usefully lay down some hard-and-fast rules of evidence that must be observed before we accept cause and effect. None of my nine viewpoints can bring indisputable evidence for or against the cause-and-effect hypothesis and none can be required as a *sine qua non*.

Austin Bradford Hill, *The Environment and Disease: Association or Causation?*, 58 Proc. Royal Soc'y Med. 295, 299 (1965).[8] Bradford Hill was clear in stating that there are causal relationships that do not satisfy one or more of the considerations, and there are many relationships that are not causal that do satisfy many of the considerations.

Many of these Bradford Hill factors are riddled with exceptions. Consider the strength of association. This consideration implies that strong associations tend to be

---

[8] Available at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1898525/pdf/procrsmed00196-0010.pdf.

casual and weak associations are likely to be non-causal. Nonetheless, many strong associations are not causal. For example, the association between birth rank and Down syndrome is strong, but it is not causal. Subsequent studies disapproved the causal inference and found the association is secondary to an association between birth rank and the age of the mother, an example of confounding. Conversely, many causal associations, such as the association between cigarette smoking and cardiovascular disease, are weak. Although weak, the proposition that smoking increases the risk of cardiovascular disease is beyond dispute. Strength of association may be worth examining, but if strong associations may be noncausal and weak associations may be causal, strength of association is hardly a criterion for causal inference.

Similarly, biological plausibility often fails to determine causation. In 1860, in a prizewinning medical essay, David Cheever described the absurdity of ascribing causation to factors that were merely associated but lacking in plausibility. David W. Cheever, *The value and the fallacy of statistics in the observation of disease*, 63 BOSTON. MED. SURG. J. 512, 515 (1861).[9] He offered an illustration:

> It could be no more ridiculous for the stranger who passed the night in the steerage of an emigrant ship to ascribe the typhus, which he there contracted, to the vermin with which the bodies of the sick might be infested.

---

[9] Available at https://collections.nlm.nih.gov/catalog/nlm:nlmuid-27720520R-bk.

*Id.* In 1861, it was not yet known that lice were the vector that spread typhus. The point is that plausibility is best seen in hindsight. Cheever's example lives on as an illustration of how plausibility can be a vague and misleading consideration, and why it should not be on a checklist of criteria for causal inference.

The process of causal inference has been studied and debated by both scientists and philosophers for centuries. Yet, knowledge about the process of causal inference has evolved to the point where several precepts are now widely accepted.

First, the strongest evidence for a causal conclusion comes not from checklists of considerations, but from crucial studies that test competing theories in a head-to-head comparison that eliminates one or more theories from consideration. The philosopher Karl Popper, perhaps the most influential philosopher of science, described this process as a Darwinian competition between theories—survival of the fittest theory. For example, when an epidemic of fatal toxic shock syndrome occurred in the 1980s, it was evidently related to a new variety of highly absorbent tampons. One theory suggested that the new tampons contained a toxin that caused toxic shock. An alternative theory suggested that the new tampons, which did not require changing as frequently as other tampons, served as a culture medium for toxic bacteria. The first theory predicts that women who change tampons more frequently would be exposed to more toxin and be at higher risk. The alternative theory predicts that changing tampons more frequently would lead to lower risk, because fresh

tampons would be introduced before the bacterial growth becomes problematic. Studies of risk according to frequency of changing tampons refuted the first theory, leaving the second as the surviving theory. This process of crucial tests illustrates the logic of causal inference. It does not involve criteria of statistical significance or checklists of vague conditions.

Second, causal connections cannot be proven, only disproven. This principle dates to the work of David Hume in the 18th century. Unlike propositions in mathematics, where logical consequences can be proven to follow from a given set of starting assumptions, in empirical science no statement about nature can be proven in a logical sense. Instead, what is taken to be scientific knowledge amounts to a set of consistent, thoroughly tested, informed guesses about how nature works. Thus, the "laws" of motion described by Galileo and Newton were accepted for centuries as an accurate description of how forces and objects interact. These laws, however, had to be revised in the 20th century when it was discovered that they did not apply in more extreme circumstances, and Newtonian mechanics was superseded by a more general theory known as relativity theory. Any statement about nature must be considered tentative, at best a theory that fits existing facts better than alternatives but nonetheless still a theory, not capable of being proven.

Third, no checklist of conditions or criteria constitutes a litmus test for a causal relation. Since the time of David Hume, philosophers have debated the methodology

for causal inference, attempting to find a logical basis for inference about empirical phenomena. This philosophical wave culminated in the work of the logical positivists in the early 20th century, who attempted a formal logic for causal inference. Popper refuted the logical positivists' attempts to circumvent Hume's skepticism about proving causal connection. Popper's philosophy of "conjecture and refutation" accepted the reality that causal connections are based on theories that cannot be proven. But Popper argued that they can be disproven when the theories are shown to make predictions that are wrong. Therefore, according to Popper, the way that science progresses is by posing theories—conjectures—and attempting to refute them. Famously, Einstein predicted that the position of stars in the sky that were aligned with the sun would change during a total eclipse, shifted by the gravity of the sun curving nearby space. Einstein claimed that if that did not happen, relativity theory would be proven incorrect. When it was verified during a subsequent total eclipse, it did not prove relativity theory to be correct, but the theory had survived a crucial test.

Ultimately, the Court should accept Bradford Hill's advice. An informal list of considerations should not substitute for rigorous scientific evaluation, examining competing explanations for observed associations. Ultimately, causation is a judgment based on information about what other explanations might account for an association.

## CONCLUSION

Statistical significance testing and the application of the Bradford Hill considerations are two approaches that are commonly but incorrectly used as a platform for causal inference. The reason both approaches are popular may be because both approaches represent a tool that absolves the investigator from the difficult job of judging a causal theory against non-causal alternatives. These methods allow mindless labels ("statistically significant", "strong association", "plausible association", etc.) and a checklist mentality to supplant the reasoning and judgment needed for scientific inference. It is considerably easier for someone to say that I have a statistically significant association that meets all the Bradford Hill criteria, and therefore must be causal, than to consider actual noncausal alternative explanations and examine how these stand up against causal explanations in crucial tests. Similarly, it is easier to dismiss an association as not statistically significant and therefore nonexistent than to accept the reality that even strong, causal associations will appear to be not statistically significant in various settings, or in subgroups.

## CERTIFICATE OF COMPLIANCE

Pursuant to Fed.R.App.P. 32(g), I certify that this brief complies with the type-volume limitations. This brief contains a total of 4,305 words.

Respectfully submitted,

s/Benjamin James Stevenson
**Benjamin James Stevenson**
Stevenson Legal, PLLC
919 Panferio Drive
Pensacola Beach, FL 32561
T. 702.306.6708
bjs@stevenson-legal.com

*Counsel for Amicus Curiae Rothman*